Decision-theoretic defaults

David Poole
Department of Computer Science,
University of British Columbia,
Vancouver, B.C., Canada V6T 1Z2
email: poole@cs.ubc.ca

Abstract

This paper considers defaults as summaries of decision-theoretic deliberations. We investigate the idea that the default $e \rightarrow a$ means that a is the optimal action based on all we know (contingently) being e. It is shown how this notion of a default is nonmonotonic and has a preference for more specific defaults. It has the advantage of defaults can, in principle, be derived from lower level concepts. We thus have a rational basis for determining whether a default is correct or not. One special case considered is where the action is whether to accept some proposition as true, accept it as false or neither. This is needed to allow for conclusions to be used as premises in other defaults. It is shown that when the gain in utility of accepting a proposition depends only only on the truth of the proposition, then the acceptance of q based on evidence e depends only on whether $P(q|e)$ exceeds a threshold that is a function of the utilities for accepting q. We also give a bound on the loss (in utility terms) of using an accepted proposition in another derivation.

1 Introduction

In AI, formal default reasoning started off as a spin off from logic [Reiter, 1980; McCarthy, 1980; McDermott and Doyle, 1980]. Logic is a normative theory of correct reasoning; the hope was that by adding in a "nonmonotonic" component, a normative theory of reasoning where we jump to conclusions could be derived. Probability theory on the other hand started off as a normative theory of reasoning under uncertainty [Jeffreys, 1961], but is very quantitative in nature. Recently qualitative versions of probability theory have been proposed for default reasoning [Pearl, 1989; Neufeld, 1989; Bacchus, 1989]. One problem with all of these proposals (with some notable exceptions [Neufeld, 1989; Bacchus, 1989]) is that we cannot "take the semantics seriously"; there is no way to use the semantics to decide whether some default is correct or not. When we do take the semantics seriously it is not so obvious that the default statements say what we actually want to say.

1.1 Defaults and utilities

What one is prepared to say "yes" to depends on both utility and probabilistic information (Doyle [1989] argues this most strongly; see section 6.1).

If one is playing a game like "trivial pursuit" (where there is no penalty for saying something wrong over the penalty for saying nothing), it is better to have a wild guess at something than to say nothing. If one is in court acting as an expert witness, then one should only say what one is sure of; witnesses don't want to be caught out and have their credibility ruined. What one assents to, and so what defaults one uses, is very dependent on the situation and the utilities involved.

If one is in a closed room full of mixes of birds and someone opens up some windows high up in the room, then what one believes about the prototypical bird in the room changes as the proportion of the flying birds in the room changes[1]. At the start we may believe that the prototypical default bird in the room flies, but as the population of the birds change, after half an hour we may believe that the prototypical bird in the room does not fly. Thus probabilistic information (information about proportions of populations with certain properties) does affect the defaults we make.

In this paper we consider a formulation of defaults that takes probability and utility into consideration.

1.2 The Proposal

Other people have observed that utilities have something to do with default reasoning [Shoham, 1987; Loui, 1990; Doyle, 1989; Kadie, 1988]. In this paper we take this relationship seriously and treat defaults as decision theory summaries.

A default $e \rightarrow_A a$ means that a is the best decision out of those decisions in A if all you know is e. Note that the conclusion of a default is an action, and not a proposition as in most default frameworks (but see section 4).

The default "if you are in Vancouver in November, carry an umbrella", is of this type of a default that has an action as a conclusion and propositions as premises. This is represented as

$$in\_Vancouver \wedge is\_November \rightarrow carry\_umbrella$$

[1] This example is due to Alan Mackworth (personal communication)

The main feature of this framework are:

- We develop a meaning for defaults and inherit a calculus (albeit very weak) for reasoning with these defaults that is sound with respect to the semantics.
- We can take the semantics seriously, and argue whether or not some default is true or not. Moreover, it can be argued that these default statements are the sort of statements that correspond to everyday defaults.
- This is useful in its own right as a summary of what actions should be taken based on certain evidence. For example, in some implementations of influence diagrams (those that evaluate the diagram independently of any particular observations) [Shachter, 1986], the output is a contingency table of the output for all tuples of possible observations. One of the motivations for this paper was to allow for a more compact representation of the decisions based on different combinations of observations.
- Building on a decision theoretic base, we develop the notion of approximate reasoning, where we can have a measure on the cost of making a mistake. This is useful when we want to develop a theory of fast, but approximate reasoning.

In section 4 we consider the special case of where the actions are to accept some proposition, to accept its negation or to accept neither. This is special as it allows us to use the conclusion of the default as a premise for more inference. We analyse the possible costs of making this unsound but often reasonable rule.

2 Background

2.1 Probability

We use a standard definition of conditional Bayesian probability (e.g., [Jeffreys, 1961]), where $P(\alpha|\beta)$ is a function from two propositions into the interval $[0,1]$, where $\beta \not\equiv false$. We use the formulation based on the three axioms:

1. $P(x|x) = 1$
2. $P(\neg x|y) = 1 - P(x|y)$
3. $P(x \wedge y|z) = P(x|y \wedge z) \times P(y|z)$

The following lemma can be easily proven from the axioms and is used in this paper:

Lemma 2.1 $P(x|z) = P(x|y \wedge z) \times P(y|z) + P(x|\neg y \wedge z) \times P(\neg y|z)$

We use the symbol $\Rightarrow$ for normal logical (material) implication.

Lemma 2.2 if $y \Rightarrow z$ then $P(x|y \wedge z) = P(x|y)$.

2.2 Classical decision theory

Under classical Bayesian decision theory (e.g., [Raiffa, 1968]), we assume that there is a subjective utility function, $\mu(a,w)$ of the utility of action a if the world is w.

The expected utility of action a given evidence e, $\mathcal{E}(a,e)$ is given by

$$\mathcal{E}(a,e) = \sum_w \mu(a,w) \times P(w|e)$$

This is the utility of a averaged over all possible worlds, weighted by their probability.

3 Decision-theoretic defaults

If e is a formula in the propositional calculus, and $A = \{a_1, a_2, ...\}$ is a set of possible alternate actions (the possible actions being primitive), and $a \in A$, we write

$$e \rightarrow_A a$$

if

$$\mathcal{E}(a,e) = \max_{a_i \in A} \mathcal{E}(a_i, e).$$

In other words, $e \rightarrow_A a$ if, given all that we know (contingently — see [Poole, 1991]) is e, a is the action in A that maximises expected utility.

3.1 Nonmonotonicity

The following example shows how the meaning of defaults can allow us to derive defaults from lower level constructs. Note that, in general the user would not provide the probability and utility, but only provide the default. Because we have a formal definition of the truth of a default, we can argue about whether some default is reasonable (based on whether the underlying probabilities and defaults are reasonable). This example also shows the notion of defaults is nonmonotonic and shows how we have a preference for more specific defaults.

Example 3.1 Suppose we have the possible actions

$$A = \{say\_flies, say\_not\_flies, say\_nothing\}$$

and the following underlying utility and probability information:

$$\begin{aligned}
\mu(say_flies, flies) &= 100 \\
\mu(say_flies, \neg flies) &= -200 \\
\mu(say_not_flies, flies) &= -200 \\
\mu(say_not_flies, \neg flies) &= 100 \\
\mu(say_nothing, flies) &= 0 \\
\mu(say_nothing, \neg flies) &= 0 \\
P(flies|bird) &= 0.9 \\
P(flies|emu) &= 0.001 \\
P(bird|emu) &= 1
\end{aligned}$$

Given $bird$ we can derive the following expected utilities

$$\begin{aligned}
&\mathcal{E}(say_flies, bird) \\
&= \mu(say_flies, flies) \times P(flies|bird) \\
&\quad + \mu(say_flies, \neg flies) \times P(\neg flies|bird) \\
&= 100 \times 0.9 - 200 \times 0.1 \\
&= 70 \\
&\mathcal{E}(say_not_flies, bird) \\
&= \mu(say_not_flies, flies) \times P(flies|bird) \\
&\quad + \mu(say_not_flies, \neg flies) \times P(\neg flies|bird) \\
&= -200 \times 0.9 + 100 \times 0.1 \\
&= -170 \\
&\mathcal{E}(say_nothing, bird)
\end{aligned}$$

$$
\begin{aligned}
&= \mu(say\_nothing, flies) \times P(flies|bird) \\
&\quad +\mu(say\_nothing, \neg flies) \times P(\neg flies|bird) \\
&= 0
\end{aligned}
$$

Thus we can derive the default

$$bird \rightarrow_A say\_flies$$

Similarly if we were given *emu*, we can compute the expected utility as:

$$
\begin{aligned}
\mathcal{E}(say\_flies, emu) &= -199.9 \\
\mathcal{E}(say\_not\_flies, emu) &= 99.7 \\
\mathcal{E}(say\_nothing, emu) &= 0
\end{aligned}
$$

Thus we can derive the default

$$emu \rightarrow_A say\_not\_flies$$

If we are given $bird \wedge emu$ we can use lemma 2.2 to show that the same deviation works as when we are given just *emu* and so we have:

$$bird \wedge emu \rightarrow_A say\_not\_flies$$

There are two things that can be derived from this example

nonmonotonicity When we learnt new information, namely that the individual was an emu as well as a bird, we no longer derived the conclusion say_flies, but rather derived the conclusion say_not_flies. We thus change our minds when presented with different information.

specificity If we know that emus are birds, when we have both *emu* and *bird*, we make the same conclusion that we would using just *emu*. This preference for more specific defaults is true in general (section 3.2).

3.2 Specificity

One of the features of defaults that is important is the fact that more specific defaults should over-ride more general defaults. If we have $x \Rightarrow y$, then knowing x is more specific knowledge than knowing y. When we have the defaults $x \rightarrow_A a$ and $y \rightarrow_A b$, then when given $x \wedge y$ we should conclude, by specificity, a. The following proposition establishes this:

Proposition 3.2 If $x \Rightarrow y$ and $x \rightarrow_A a$ then $x \wedge y \rightarrow_A a$.

The proof of this and other propositions appears in Appendix A.

This result should not be too surprising, as the preference for more specific knowledge is common to the probabilistic formulations of defaults (see [Pearl, 1989]), as opposed to the logical formulations of defaults.

3.3 Ignoring Irrelevance

If we have some condition c such that we make the same decision whether or not c is true, then we will make that decision even if we did not know the truth of c.

Proposition 3.3 The following is valid inference:

$$\frac{e \wedge c \rightarrow_A a \quad e \wedge \neg c \rightarrow_A a}{e \rightarrow_A a}$$

Thus if would make the same decision if c is true or false, then we can ignore the value of c in our defaults. What is important here is that we can derive consequences of our defaults based on the underlying definition.

3.4 Disjunction

We cannot do arbitrary reasoning by cases. For example the disjunction rule of Pearl [1989]

$$\frac{e_1 \rightarrow_A a \quad e_2 \rightarrow_A a}{e_1 \vee e_2 \rightarrow_A a}$$

is not valid in general. It is however valid when e_1 and e_2 either imply each other or are inconsistent. Accepting this rule would lead us to Simpson's paradox [Neufeld and Horton, 1990].

3.5 Restricting the choices

Sometimes we may have fewer choices to make than at other times. The following proposition shows that if we do not eliminate the best choice, we can restrict the choices available without affecting the default.

Proposition 3.4 If $e \rightarrow_A a$ and $B \subseteq A$ such that $a \in B$ then $e \rightarrow_B a$.

We can use the following lemma to split the set of possible alternatives.

Proposition 3.5 If $e \rightarrow_A a$ and $e \rightarrow_B b$ then either $e \rightarrow_{A \cup B} a$ or $e \rightarrow_{A \cup B} b$

In the rest of the paper we assume that the set of choices of actions is fixed, and omit the subscript to $\rightarrow$.

4 Acceptance Assumption

The preceding section considered the case when the conclusion of a default was an action. Often in default reasoning, we want to use a default to conclude that some proposition is true, and then use that proposition in further reasoning.

Jon Doyle has previously propounded the idea that we expand on:

> "... we wish to use rationality as a standard for adopting assumptions by saying that an assumption should be adopted if the expected utility of holding it exceeds the expected utility of not holding it." [Doyle, 1989, p. 5]

In this section we show how to relate the action that is a conclusion of a default to the acceptance of the truth of a proposition. We consider the acceptance of a proposition as a decision like another decision. For a proposition z there are three alternate decisions that could be made:

- z^t is the decision to accept proposition z as true.
- z^f is the decision to accept z as false.
- z^u is the decision to neither accept z nor $\neg z$.

For each proposition we make the decision of whether to accept it as true, to accept it as false or to make no commitment.

Example 4.1 We write the decision to accept *flies* as true if the individual under consideration is a bird as

$$bird \rightarrow flies^t$$

This would correspond to the default in example 3.1, but the action is to accept the proposition *flies*, rather than the action to say something. Similarly the default that injured birds do not fly, can be written:

$$bird \wedge injured \rightarrow flies^f$$

This says that if all you know about some individual is that the individual is an injured bird, that it is better to assume that it does not fly, than being uncommitted about the flying ability of the individual or assuming that it does fly.

We also allow for a default to conclude that we should not assume anything about the flying ability of young birds:

$$bird \wedge young \rightarrow flies^u$$

Note that we only have a two valued logic (classical probability theory is based on every proposition being true or false in each possible world [Jeffreys, 1961]), but we have three possible actions that we can do with respect to a proposition. We assume here that it is never a good policy to assume a proposition and its negation.

Assumption 4.2 We assume the following inequalities:

$$\mu(z^f|z) < \mu(z^u|z) < \mu(z^t|z)$$

$$\mu(z^t|\neg z) < \mu(z^u|\neg z) < \mu(z^f|\neg z)$$

That is it is better to guess correctly than to be non-committal. And it is better to be non-committal than guessing wrongly[2].

The second assumption that we make is that the utilities of different propositions are in some sense independent. We can treat the gain in accepting proposition z as not affected by the truth of other propositions.

Assumption 4.3 The change in utility of accepting z or accepting $\neg z$ or accepting neither in a world w depends only on the truth of z in w.

That is, if w_1 and w_2 are two worlds that agree on the truth of z then

$$\mu(z^r, w_1) - \mu(z^s, w_1) = \mu(z^r, w_2) - \mu(z^s, w_2)$$

where $\{r, s\} \subset \{t, u, f\}$.

This assumption means that we only have to consider the gain in making the correct decision and the loss in making an incorrect decision.

Definition 4.4 If p is an atomic proposition we use the following notational schema where r and s denote different elements of $\{t, f, u\}$, and σ is a sign (one of $+$ or $\neg$) such that σp is p if σ is $+$ and σp is $\neg p$ is σ is $\neg$. We define

$${}^r\Delta^s(\sigma p) = \mu(p^s, w) - \mu(p^r, w)$$

where w is a world in which σp is true. This schema, representing 12 different formulae, denotes the change in utility made what changing our action from r to s.

[2] Analogous results to the ones below hold when the above constraints are violated — the arithmetic is slightly changed.

For example, ${}^f\Delta^t(z)$ is $\mu(z^t, w) - \mu(z^f, w)$, where z is true in w, which is the utility gained when we decide to commit to z over committing to $\neg z$ given that z is true. ${}^u\Delta^f(\neg z)$ is $\mu(z^f, w) - \mu(z^u, w)$, where z is false in w, which is the utility gained when we decide to commit to $\neg z$ over not committing to the truth of z given that z is false.

Under assumption 4.2, ${}^f\Delta^t(z)$, ${}^f\Delta^u(z)$ and ${}^u\Delta^t(z)$ are all positive. These all consist of the gain made by making a better guess given that z is true. Note also that

$${}^f\Delta^t(z) = {}^f\Delta^u(z) + {}^u\Delta^t(z).$$

Thus ${}^t\Delta^f(\neg z)$, ${}^u\Delta^f(\neg z)$ and ${}^t\Delta^u(\neg z)$ are all positive. All of the others are negative, using the equality

$${}^r\Delta^s(\sigma z) = -{}^s\Delta^r(\sigma z)$$

Lemma 4.5 For s and r each being one of t, u or f, the following holds:

$$\mathcal{E}(z^s, x) - \mathcal{E}(z^r, x) = {}^r\Delta^s(z) \times P(z|x) + {}^r\Delta^s(\neg z) \times P(\neg z|x)$$

4.1 Characterization of defaults

In this section we analyse when we can conclude a default based on the assumptions in the previous section. We first consider when one decision should be made over another decision.

Lemma 4.6

$$\mathcal{E}(z^t, x) \geq \mathcal{E}(z^f, x)$$

if and only if

$$P(z|x) \geq \frac{{}^t\Delta^f(\neg z)}{{}^f\Delta^t(z) + {}^t\Delta^f(\neg z)}$$

Lemma 4.7

$$\mathcal{E}(z^t, x) \geq \mathcal{E}(z^u, x)$$

if and only if

$$P(z|x) \geq \frac{{}^t\Delta^u(\neg z)}{{}^u\Delta^t(z) + {}^t\Delta^u(\neg z)}$$

The following theorem is a direct corollary of lemmata 4.6 and 4.7.

Theorem 4.8 $x \rightarrow z^t$ if and only if

$$P(z|x) \geq \max\left(\frac{{}^t\Delta^f(\neg z)}{{}^f\Delta^t(z) + {}^t\Delta^f(\neg z)}, \frac{{}^t\Delta^u(\neg z)}{{}^u\Delta^t(z) + {}^t\Delta^u(\neg z)}\right)$$

What is important to notice here is that the decision to accept z based on x is determined completely by a threshold on the probability $P(z|x)$ and that the threshold is a function of the utilities of the acceptance of z.

We can carry out a similar analysis of

Lemma 4.9

$$\mathcal{E}(z^u, x) \geq \mathcal{E}(z^f, x)$$

if and only if

$$P(z|x) \geq \frac{{}^u\Delta^f(\neg z)}{{}^f\Delta^u(z) + {}^u\Delta^f(\neg z)}$$

The following theorem is a direct corollary of lemmata 4.6, 4.7 and 4.9.

Theorem 4.10 $x \rightarrow z^t$ iff

$$P(z|x) \geq \max\left(\frac{{}^t\Delta^f(\neg z)}{{}^f\Delta^t(z) + {}^t\Delta^f(\neg z)}, \frac{{}^t\Delta^u(\neg z)}{{}^u\Delta^t(z) + {}^t\Delta^u(\neg z)}\right)$$

$x \rightarrow z^f$ iff

$$P(z|x) \leq \min\left(\frac{{}^t\Delta^f(\neg z)}{{}^f\Delta^t(z) + {}^t\Delta^f(\neg z)}, \frac{{}^u\Delta^f(\neg z)}{{}^f\Delta^u(z) + {}^u\Delta^f(\neg z)}\right)$$

$x \rightarrow z^u$ iff $P(z|x)$ is between these two values.

Example 4.11 Suppose we have the following utilities

$$\begin{aligned}\mu(p^u, p) = \mu(p^u, \neg p) &= 0\\ \mu(p^t, p) = \mu(p^f, \neg p) &= a\\ \mu(p^t, \neg p) = \mu(p^f, p) &= -b\end{aligned}$$

a is the prize we get for guessing right. b is the price we pay if we are wrong; both a and b are positive.

We have:

$$\begin{aligned}\frac{{}^t\Delta^f(\neg z)}{{}^f\Delta^t(z) + {}^t\Delta^f(\neg z)} &= 0.5\\ \frac{{}^t\Delta^u(\neg z)}{{}^u\Delta^t(z) + {}^t\Delta^u(\neg z)} &= \frac{b}{a+b}\\ \frac{{}^u\Delta^f(\neg z)}{{}^f\Delta^u(z) + {}^u\Delta^f(\neg z)} &= \frac{a}{a+b}\end{aligned}$$

We have the following cases of acceptance:

$a > b$

$$\begin{aligned}q \rightarrow p^t \quad &\text{if} \quad P(p|q) \geq \frac{1}{2}\\ q \rightarrow p^f \quad &\text{if} \quad P(p|q) \leq \frac{1}{2}\end{aligned}$$

Here we would never decide on p^u. We would expect to lose by being noncommittal.

$a = b$

$$\begin{aligned}q \rightarrow p^t \quad &\text{if} \quad P(p|q) \geq \frac{1}{2}\\ q \rightarrow p^u \quad &\text{if} \quad P(p|q) = \frac{1}{2}\\ q \rightarrow p^f \quad &\text{if} \quad P(p|q) \leq \frac{1}{2}\end{aligned}$$

Here, when $P(p|q) = \frac{1}{2}$, it doesn't matter which decision we make. They all have the same expected utility.

$a < b$

$$\begin{aligned}q \rightarrow p^t \quad &\text{if} \quad P(p|q) \geq \frac{b}{a+b}\\ q \rightarrow p^f \quad &\text{if} \quad P(p|q) \leq \frac{a}{a+b}\\ q \rightarrow p^u \quad &\text{if} \quad \frac{a}{a+b} \leq P(p|q) \leq \frac{b}{a+b}\end{aligned}$$

If we are very conservative we would expect that $b \gg a$. In this case we have

$$\frac{b}{a+b} \approx 1$$

The algebra of thresholding probability here is the same as the system of Bacchus [1989], but where the actual value of the threshold depends on the utilities associated with the acceptance of the conclusion of the default.

5 Approximate Reasoning

One of the features of utility-based approach to default reasoning is the ability to have a notion of the cost of getting a wrong answer. We can thus talk precisely about a tradeoff of accuracy, and consider the cost of making assumptions. Consider the following rule (called "contraction" [Pearl, 1989]):

$$\frac{\begin{array}{c}x \rightarrow y^t\\ x \wedge y \rightarrow z^t\end{array}}{x \rightarrow z^t}$$

This rule says that if we can conclude y, and use y to conclude z then we can conclude z without using y. This rule says that we can use derived conclusions as lemmata for other conclusions. This is not a valid rule of inference in the decision-theoretic defaults [Bacchus, 1989]. This is because we do not know that y is true we have only decided that we should make it true.

Pearl [1989] argues that ϵ-semantics (in which contraction is a valid rule) is an idealisation. One of the main advantages of the decision-theoretic defaults is that we can measure the cost of our idealisation. With the decision-theoretic defaults we can consider how much we can lose by applying the above rule.

Proposition 5.1 The maximum that we can lose by following the above rule is

$$(1 - th(y^t)) \times {}^t\Delta^f(\neg z)$$

where $th(y^t)$ is the threshold for accepting y, which is

$$th(y^t) = \max\left(\frac{{}^t\Delta^f(\neg z)}{{}^f\Delta^t(z) + {}^t\Delta^f(\neg z)}, \frac{{}^t\Delta^u(\neg z)}{{}^u\Delta^t(z) + {}^t\Delta^u(\neg z)}\right)$$

Example 5.2 Using the utilities of example 4.11, we find that the maximum we can lose is

$$\begin{aligned}&(1 - th(y^t)) \times {}^t\Delta^f(\neg z)\\ =\ &\frac{a}{a+b} \times (a+b)\\ =\ &a\end{aligned}$$

Even if we are extremely conservative and have a large b value, the conservatism in the acceptance of y means that we cannot lose much when we accept z.

6 Comparison with other proposals

6.1 Doyle

Doyle has also considered the role of utility and probability in default reasoning [Doyle, 1989]. This paper can be seen as following in the pioneering steps of [Doyle, 1989]

in incorporating rationality into reasoning. We go into much more detail in one case of the general framework outlined by Doyle.

Doyle [1990] motivates his rational belief revision in economic terms. However, unlike the defaults in this paper, the object level statements are not statements of preference in a utility sense. The utility is to suggest alternate definitions of belief revision. I would argue that the notion of utility of beliefs should be logically prior to the notion of rational belief revision. Once we have a notion of the utility of belief, we should be able to use this to develop a notion of rational belief revision.

Other work of Doyle [1985; 1989] has considered the problem of default reasoning as a problem of group decision making, and used the theory of group decision making for default reasoning. The group decision making and the individual decision making used in this paper are not incompatible (unless we want to claim they are the same [Doyle and Wellman, 1989]), and so these approaches should be seen as complementary to the approach propounded here.

6.2 Shoham

Shoham [1987] has argued that we should take probabilities and utilities into account when considering defaults. Here we take this suggestion seriously and consider the normative theory of decision making as a starting point. He instead develops a general framework of nonmonotonic reasoning based on ordering of interpretations. The system propounded here cannot be simply put into the framework developed by Shoham (one of the reasons is that we have automatic specificity, which one can show cannot be in any system that treats all logically equivalent formulae as equivalent [Poole, 1991]).

6.3 Loui

Loui [1990] has also proposed a mix between decision theory and defeasible reasoning. He has, however, suggested the opposite mix, namely using a form of defeasible reasoning for decision making. His motivation is very different to the motivation of this paper; it is an intriguing idea to consider whether the default system propounded here could be used as the basis for the argument system in Loui's proposal.

6.4 Bacchus

Bacchus [1989] has investigated the logic of thresholding conditional probability. All of the results of his theory can be transferred to the system in this paper. We complement Bacchus' work in that we show how straightforward decision-theoretic concerns lead us to thresholding probability.

Rather than having a constant threshold for acceptance, we have a different threshold for each proposition. While this is not inconsistent with Bacchus's results, it is interesting that we can determine exactly what the threshold should depend on. This is because we can answer the question of where the thresholds come from.

In Bacchus's system, all one can say about such rules as contraction (section 5) is that they are unsound with respect to the thresholding semantics. In the system outlined in this paper we can answer the question of how much we can lose by using these idealised rules of inference. and look at the utility of using conclusions, even if they may be mistaken.

7 Conclusion

In this paper we considered a simple idea; namely that defaults provide summaries of possible decisions that has already taken utilities and probabilities into considerations. This allows for a definition of default for which we can take the meaning seriously. I would argue that the default "birds fly" really means that if all you know about some individual is a bird, then it is good policy to assume that the individual can fly. Rather than using decision theory directly for nonmonotonic reasoning [Kadie, 1988], this paper has explored only having the summaries of good decisions as defaults.

The main result was to show that under the assumption that the utility of the choice of whether to accept a proposition depends only on the truth of the proposition (assumption 4.3), the acceptance depends on thresholds of conditional probability. Thus we get to the same system that Bacchus [1989] proposed. We have the advantage that we can derive the threshold for acceptance from utility considerations. This is one of the few proposals that can use the idea of the cost of an incorrect conclusion.

The resulting calculus is very weak. Further work can be carried out in incorporating independence assumptions, and in making assumption 4.3 more realistic. Assumption 4.3 is interesting as an idealisation, but is not practical. In practice the importance of a piece of information critically depends on what other information is true.

A Proofs

Proposition 3.2 If $x \Rightarrow y$ and $x \rightarrow a$ then $x \wedge y \rightarrow a$.

Proof: If $x \Rightarrow y$ then $P(y|x) = 1$.

$$\begin{aligned}
&\mathcal{E}(a_i, x \wedge y) \\
&\quad = \sum_w \mu(a_i, w) \times P(w|x \wedge y) \\
&\quad = \sum_w \mu(a_i, w) \times P(w|x) \text{ (by lemma 2.2)} \\
&\quad = \mathcal{E}(a_i, x)
\end{aligned}$$

The result follows immediately. □

Proposition 3.3 The following is valid inference:

$$\frac{e \wedge c \rightarrow_A a \quad e \wedge \neg c \rightarrow_A a}{e \rightarrow_A a}$$

Proof: Using lemma 2.1, we have

$$\begin{aligned}
&P(w|e) \\
&\quad = P(w|e \wedge c) \times P(c|e) \\
&\qquad + P(w|e \wedge \neg c) \times P(\neg c|e) \\
&\mathcal{E}(a_i, e)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_w \mu(a_i, w) \times P(w|e) \\
&= P(c|e) \sum_w \mu(a_i, w) \times P(w|e \wedge c) \\
&\quad + P(\neg c|e) \sum_w \mu(a_i, w) \times P(w|e \wedge \neg c) \\
&= P(c|e)\mathcal{E}(a_i, e \wedge c) + P(\neg c|e)\mathcal{E}(a_i, e \wedge \neg c)
\end{aligned}
$$

We know $e \wedge c \rightarrow a$ and $e \wedge \neg c \rightarrow a$, so for each $a_i \in A$,

$$
\begin{aligned}
\text{given } \mathcal{E}(a, e \wedge c) &\geq \mathcal{E}(a_i, e \wedge c) \\
\text{and } \mathcal{E}(a, e \wedge \neg c) &\geq \mathcal{E}(a_i, e \wedge \neg c) \\
\text{then } P(c|e)\mathcal{E}(a, e \wedge c) + & \\
P(\neg c|e)\mathcal{E}(a, e \wedge \neg c) &\geq P(c|e)\mathcal{E}(a_i, e \wedge c) + \\
& \quad P(\neg c|e)\mathcal{E}(a_i, e \wedge \neg c) \\
\text{so } \mathcal{E}(a, e) &\geq \mathcal{E}(a_i, e)
\end{aligned}
$$

□

Lemma 4.5 For s and r each being one of t, u or f, the following holds:

$$\mathcal{E}(z^s, x) - \mathcal{E}(z^r, x) = {}^r\Delta^s(z) \times P(z|x) + {}^r\Delta^s(\neg z) \times P(\neg z|x)$$

Proof:

$$
\begin{aligned}
&\mathcal{E}(z^s, x) - \mathcal{E}(z^r, x) \\
&= \sum_w \mu(z^s, w) \times P(w|x) - \sum_w \mu(z^r, w) \times P(w|x) \\
&= \sum_{w:z \text{ true in } w} (\mu(z^s, w) - \mu(z^r, w)) \times P(w|x) \\
&\quad + \sum_{w:z \text{ false in } w} (\mu(z^s, w) - \mu(z^r, w)) \times P(w|x) \\
&= {}^r\Delta^s(z) \sum_{w:z \text{ true in } w} P(w|x) \\
&\quad + {}^r\Delta^s(\neg z) \sum_{w:z \text{ false in } w} P(w|x) \\
&= {}^r\Delta^s(z) \times P(z|x) + {}^r\Delta^s(z) \times P(\neg z|x)
\end{aligned}
$$

□

Lemma 4.6

$$\mathcal{E}(z^t, x) \geq \mathcal{E}(z^f, x)$$

if and only if

$$P(z|x) \geq \frac{{}^t\Delta^f(\neg z)}{{}^f\Delta^t(z) + {}^t\Delta^f(\neg z)}$$

Proof: The following sequence of inequalities are all equivalent:

$$
\begin{aligned}
\mathcal{E}(z^t, x) &\geq \mathcal{E}(z^f, x) \\
\mathcal{E}(z^t, x) - \mathcal{E}(z^f, x) &\geq 0 \\
{}^f\Delta^t(z) \times P(z|x) - {}^t\Delta^f(\neg z) \times P(\neg z|x) &\geq 0 \\
{}^f\Delta^t(z) \times P(z|x) &\geq {}^t\Delta^f(\neg z) \times (1 - P(z|x)) \\
({}^f\Delta^t(z) + {}^t\Delta^f(\neg z)) \times P(z|x) &\geq {}^t\Delta^f(\neg z) \\
P(z|x) &\geq \frac{{}^t\Delta^f(\neg z)}{{}^f\Delta^t(z) + {}^t\Delta^f(\neg z)}
\end{aligned}
$$

□

The proofs of lemmata 4.7 and 4.9 are analogous to the proof of lemma 4.6, and are omitted.

Proposition 5.1 The maximum that we can lose by following the rule of contraction is

$$(1 - th(y^t)) \times {}^t\Delta^f(\neg z)$$

where $th(y^t)$ is the threshold for accepting y, which is

$$th(y^t) = \max\left(\frac{{}^t\Delta^f(\neg z)}{{}^f\Delta^t(z) + {}^t\Delta^f(\neg z)}, \frac{{}^t\Delta^u(\neg z)}{{}^u\Delta^t(z) + {}^t\Delta^u(\neg z)}\right)$$

Proof: Suppose we have $x \rightarrow y^t$ and $x \wedge y \rightarrow z^t$. The maximum we can lose by using the rule $x \rightarrow z^t$ is given by how much we would gain by doing one of the other two actions. This is

$$max\left(\mathcal{E}(z^u, x) - \mathcal{E}(z^t, x), \mathcal{E}(z^f, x) - \mathcal{E}(z^t, x)\right)$$

For s being either of u or f, we can derive

$$
\begin{aligned}
&\mathcal{E}(z^s, x) - \mathcal{E}(z^t, x) \\
&\quad = -{}^s\Delta^t(z) \times P(z|x) + {}^t\Delta^s(\neg z) \times (1 - P(z|x)) \\
&\quad = {}^t\Delta^s(\neg z) - ({}^s\Delta^t(z) + {}^t\Delta^s(\neg z)) \times P(z|x) \\
&P(z|x) \\
&\quad = P(z|x \wedge y) \times P(y|x) + P(z|x \wedge \neg y) \times P(\neg y|x) \\
&\quad \geq P(z|x \wedge y) \times P(y|x) \\
&\quad \geq \left(\frac{{}^t\Delta^s(\neg z)}{{}^s\Delta^t(z) + {}^t\Delta^s(\neg z)}\right) \times th(y^t) \\
&\mathcal{E}(z^s, x) - \mathcal{E}(z^t, x) \\
&\quad \leq {}^t\Delta^s(\neg z) - \left({}^s\Delta^t(z) + {}^t\Delta^s(\neg z)\right) \times \\
&\qquad \left(\frac{{}^t\Delta^s(\neg z)}{{}^s\Delta^t(z) + {}^t\Delta^s(\neg z)}\right) \times th(y^t) \\
&\quad = {}^t\Delta^s(\neg z) - {}^t\Delta^s(\neg z) \times th(y^t) \\
&\quad = {}^t\Delta^s(\neg z) \times (1 - th(y^t))
\end{aligned}
$$

So that maximum that we can lose is

$$max\left({}^t\Delta^f(\neg z) \times (1 - th(y^t)), {}^t\Delta^u(\neg z) \times (1 - th(y^t))\right)$$

which, under assumption 4.2 is ${}^t\Delta^f(\neg z) \times (1 - th(y^t))$.
□

Acknowledgements

This research was supported under NSERC grant OGPOO44121, and under Project B5 of the Institute for Robotics and Intelligent Systems.

References

[Bacchus, 1989] F. Bacchus. A modest, but semantically well founded, inheritance reasoner. In *Proc. 11th International Joint Conf. on Artificial Intelligence*, pages 1104–1109, Detroit, August 1989.

[Doyle and Wellman, 1989] J. Doyle and M. P. Wellman. Impediments to universal preference-based default theories. In R. Reiter R. J. Brachman, H. J. Levesque, editor, *Proc. First International Conf. on Principles of Knowledge Representation and Reasoning*, pages 94–102, Toronto, May 1989.

[Doyle, 1985] J. Doyle. Reasoned assumptions and pareto optimality. In *Proc. 9th International Joint Conf. on Artificial Intelligence*, pages 87–90, Los Angeles, August 1985.

[Doyle, 1989] J. Doyle. Constructive belief and rational representation. *Computational Intelligence*, 5(1):1–11, February 1989.

[Doyle, 1990] J. Doyle. Rational belief revision. In *Preprints of Third International Workshop on Nonmonotonic Reasoning*, Lake Tahoe, CA, May 1990.

[Jeffreys, 1961] H. Jeffreys. *Theory of probability.* Oxford University Press, Oxford, third edition, 1961.

[Kadie, 1988] C. M. Kadie. Rational nonmonotonic reaoning. In *Proc. Fourth Workshop on Uncertainty in Artificial Intelligence*, pages 197–204, University of Minnesota, August 1988.

[Loui, 1990] R. P. Loui. Defeasible decisions: what the proposal is and isn't. In M. Henrion et al., editor, *Uncertainty in Artificial Intelligence 5*, pages 99–116. North Holland, 1990.

[McCarthy, 1980] J. McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence*, 13(1,2):27–39, 1980.

[McDermott and Doyle, 1980] D. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13:41–72, 1980.

[Neufeld and Horton, 1990] E. Neufeld and J. D. Horton. Conditioning on disjunctive knowledge: Simpson's paradox in default logic. In M. Henrion et al., editor, *Uncertainty in Artificial Intelligence 5*, pages 117–125, 1990.

[Neufeld, 1989] E. Neufeld. Defaults and probabilities; extensions and coherence. In H. J. Levesque R. J. Brachman and R. Reiter, editors, *Proc. First International Conf. on Principles of Knowledge Representation and Reasoning*, pages 312–323, Toronto, May 1989.

[Pearl, 1989] J. Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. In H. J. Levesque R. J. Brachman and R. Reiter, editors, *Proc. First International Conf. on Principles of Knowledge Representation and Reasoning*, pages 505–516, Toronto, May 1989.

[Poole, 1991] D. Poole. The effect of knowledge on belief: conditioning, specificity and the lottery paradox in default reasoning. *Artificial Intelligence*, 49:281–307, 1991.

[Raiffa, 1968] H. Raiffa. *Decision Analysis.* Addison-Wesley, 1968.

[Reiter, 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1,2):81–132, 1980.

[Shachter, 1986] R. D. Shachter. Evaluating influence diagrams. *Operations Research*, 34(6):871–882, November-December 1986.

[Shoham, 1987] Y. Shoham. Nonmonotonic logics: Meaning and utility. In *Proc. 10th International Joint Conf. on Artificial Intelligence*, pages 388–393, Milan, August 1987.